

CLAIMS

What is claimed is:

5 1. In a communication network having a plurality of computational resources for servicing a plurality of application environments, a method for enabling resource sharing, comprising:

10 a) monitoring quality of service provided from a plurality of components coupled together in a first application environment supporting a first application;

15 b) determining whether a plurality of service level objectives are satisfied, each of said plurality of service level objectives associated with one of said plurality of components; and

20 c) optimizing the number of computational resources from said plurality of computational resources in each of said plurality of components in order to satisfy said plurality of service level objectives.

25 2. The method for enabling resource sharing as described in Claim 1, wherein a) comprises:

30 a1) determining a plurality of metrics, each of said plurality of metrics characterizing quality of service for a particular component in said plurality of components;

 a2) sending said plurality of metrics to a dynamic resource manager.

35 3. The method as described in 1, wherein said plurality of components are flexibly sized partitions of an electronic device.

4. The method for enabling resource sharing as described in Claim 2, wherein b) comprises:

b1) at said dynamic resource manager, comparing each of said plurality of metrics to an associated service level objective in said plurality of service level objectives, each of said plurality of service level objectives associated with one of said plurality of components; and

b2) determining whether each of said plurality of metrics fall within an associated interval for said associated service level objective.

5. The method for enabling resource sharing as described in Claim 4, wherein c) comprises:

sending a message to one of said plurality of components having a corresponding service level objective and corresponding interval to add at least one available computational resource from said plurality of computational resources when an associated metric exceeds said corresponding interval.

6. The method for enabling resource sharing as described in Claim 4, wherein c) comprises:

sending a message to one of said plurality of components having a corresponding service level objective and a corresponding interval to remove at least one computational resource, a removed computational resource, in said plurality of computational resources that are assigned to said one of said plurality of components when an associated metric falls short of said corresponding interval, thereby freeing up said removed computational resource for a second application supported by a second application environment in said communication network.

7. The method for enabling resource sharing as described in Claim 4, wherein c) comprises:

perform no action, if, at said dynamic resource manager, one of said plurality of metrics, that is
5 associated with a corresponding component in said plurality of components and associated with a corresponding interval for an associated service level objective, meets said corresponding interval.

10 8. The method for enabling resource sharing as described in Claim 1, wherein b) comprises:

determining each of said plurality of service level objectives.

15 9. The method for enabling resource sharing as described in Claim 2, wherein each of said plurality of metrics is a response-time metric.

20 10. In a communication network having a plurality of computational resources for supporting a plurality of application environments, a method for enabling resource sharing, comprising:

a) receiving a first response-time metric from a first component in a plurality of components that form a
25 first application environment in said plurality of application environments;

b) comparing said first response-time metric to a first service level objective associated with said first component; and

30 c) optimizing the number of computational resources in said plurality of computational resources that are assigned to said first component in order to satisfy said first service level objective.

11. The method for enabling resource sharing as described in Claim 10, further comprising:

at said first component, determining said first
5 response-time metric, said first response-time metric
characterizing quality of service for said first component;
and

sending said first response-time metric to said
dynamic resource manager.

10 12. The method for enabling resource sharing as described in Claim 10, wherein b1) further comprises:

determining whether said first response-time metric
falls within a first interval for said first service level
15 objective in order to satisfy said first service level
objective.

13. The method for enabling resource sharing as described in Claim 12, further comprising:

20 if said first response-time metric exceeds said first
interval, sending a first message to a first component
manager associated with said first component to add at
least one available computational resource in said
plurality of computational resources;

25 if said first response-time metric falls below said
first interval, sending a second message to said first
component manager to remove at least one computational
resource in said plurality of computational resources that
is assigned to said first component; and

30 perform no action, if said first response-time metric
falls within said first interval.

14. The method for enabling resource sharing as described in Claim 10, further comprising:

d) receiving a second response-time metric from a second component in said plurality of components;

5 e) comparing said second response-time metric to a second service level objective associated with said second component;

f) determining whether said second response-time metric falls within a second interval for said second
10 service level objective; and

g) optimizing the number of computational resources in said plurality of computational resources that are assigned to said second component in order to satisfy said second service level objective.

15 15. The method for enabling resource sharing as described in Claim 14, further comprising:

if said second response-time metric exceeds said second interval, sending said first message to a second
20 component manager associated with said second component;

if said second response-time metric falls below said second interval, sending said second message to said second component manager; and

perform no action, if said second response-time metric
25 falls within said second interval.

16. A communication network comprising:

a plurality of computational resources;

an application environment having a plurality of
30 network nodes coupled together;

a plurality of components in said application environment servicing an application, each of said plurality of components including at least one

computational resource from said plurality of computational resources, each of said plurality of components residing on one of said plurality of network nodes; and

a dynamic resource manager residing coupled to said application environment for optimizing the number of computational resources from said plurality of computational resources in each of said plurality of components in order to satisfy quality of service objectives for said application.

17. The communication network as described in Claim 16, further comprising:

a plurality of component managers, each of said plurality of component manager monitoring quality of service levels in one of said plurality of components, and for managing the addition and removal of computational resources in said one of said plurality of components in response to notices from said dynamic resource manager.

18. The communication network as described in Claim 16, further comprising:

a plurality of metrics, one for each of said plurality of components for measuring quality of service levels provided at each of said plurality of components, said plurality of metrics monitored by a plurality of component managers, one for each of said plurality of components.

19. The communication network as described in Claim 16, wherein said quality of service objectives further comprise:

a plurality of service level objectives, each of said plurality of service level objectives associated with one of said plurality of components, each of said plurality of

service level objectives characterizing quality of service levels for a particular component in said plurality of components.

5 20. The communication network as described in Claim 18, wherein said plurality of service level objectives further comprise:

10 a plurality of intervals, each of said plurality of intervals associated with one of said plurality of service level objectives, each of said plurality of intervals defining satisfactory ranges of metrics;

15 an upper boundary for each of said plurality of intervals, wherein metrics exceeding said upper boundary are unacceptable, requiring additional resources or computational resources at the associated node;

20 a lower boundary for each of said plurality of intervals, wherein metrics falling below said lower boundary indicate quality of service is exceeded, requiring removal of resources or computational resources at the associated node.

21. The communication network as described in Claim 18, further comprising:

25 an add message, said add message created by said dynamic resource manager when a metric for an associated component exceeds an associated interval, said add message sent from said dynamic resource manager to said associated component.

30 22. The communication network as described in Claim 18, further comprising:

 a remove message, said remove message created by said dynamic resource manager when a metric for an associated

component falls below an associated interval, said remove message sent from said dynamic resource manager to said associated component.

5 23. The communication network as described in Claim 16, further comprising:

 a second application environment supporting a second application.

 24. The communication network as described in Claim 10 18, wherein said plurality of metrics is a plurality of response-time metrics.